caBIG &
The
SNOMED CT - ENCODED CAP CANCER CHECKLISTS (SECCC)

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Richard Moldwin, MD, PhD
College of American Pathologists
SNOMED Terminology Solutions

The CAP Cancer Protocols

- The CAP publishes *cancer protocols* as a resource to pathologists in effectively delivering the information necessary to provide quality patient care.
- The "Protocols" consist of cancer case summaries ("*checklists*") accompanied by background documentation.
- These widely-used case summaries are sometimes called "synoptic reports."

Protocol for the Examination of Specimens from Patients with Primary Carcinomas of the Colon and Rectum

Well differentiated neuroendocrine tumors (carcinoid tumors) are not included.

Based on AJCC/UICC TNM, 6th edition
Protocol web posting date: _____
Protocol effective date: _____

Procedures
- Excisional Biopsy (Polypectomy)
- Local Excision (Transanal Disk Excision)
- Colectomy (Total, Partial, or Segmental Resection)
- Rectal Resection (Low Anterior Resection or Abdominoperineal Resection)

Authors

Kay Washington, MD, PhD, FCAP
→ Department of Pathology, Vanderbilt University Medical Center, Nashville, TN
Jordan Berlin, MD
→ Department of Medicine, Vanderbilt University Medical Center, Nashville, TN
Philip Branton, MD, FCAP
→ Department of Pathology, Inova Fairfax Hospital, Falls Church, VA
Lawrence J. Burgart, MD, FCAP
→ Allina Laboratories, Abbott Northwestern Hospital, Minneapolis, MN
David K. Carter, MD, FCAP
→ Department of Pathology, St. Mary's/Duluth Clinic Health System, Duluth, MN
Patrick Fitzgibbons, MD, FCAP
Department of Pathology, St. Jude Medical Center, Fullerton, CA
Wendy Frankel, MD, FCAP
Department of Pathology, Ohio State University Medical Center, Columbus, OH
John Jessup, MD
→ Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD
Sanjay Kakar, MD, FCAP
Department of Pathology, University of California San Francisco and the Veterans Affairs Medical Center, San Francisco, CA
Bruce Minsky, MD
→ Department of Radiation Oncology, University of Chicago, Chicago, IL
Raouf Nakhleh, MD, FCAP
→ Department of Pathology, St. Luke's Hospital, Jacksonville, FL
Carolyn C. Compton, MD, PhD, FCAP
→ Office of Biorepositories and Biospecimen Research, National Cancer Institute, Bethesda, MD
For the Members of the Cancer Committee, College of American Pathologists

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Sample Un-encoded Checklist

Surgical Pathology Cancer Case Summary (Checklist)

Protocol revision date: January 2005
Applies to invasive carcinomas only
Based on AJCC/UICC TNM, 6th edition

COLON AND RECTUM: Excisional Biopsy (Polypectomy)

Patient name:
Surgical pathology number:

Note: Check 1 response unless otherwise indicated.

MACROSCOPIC

Tumor Site
___ Cecum
___ Right (ascending) colon
___ Hepatic flexure
___ Transverse colon
___ Splenic flexure
___ Left (descending) colon
___ Sigmoid colon
___ Rectum
___ Not specified

© 2007
College of American Pathologists
SNOMED Terminology Solutions

SNOMED CT® ENCODING of the CANCER CHECKLISTS

SNOMED = Systematized Nomenclature of Medicine Clinical Terms

SPECIMEN TYPE *[R-00254, 371439000] Specimen type (observable entity)*
___ Polypectomy *[G-8423, 397053005] Specimen from small intestine obtained by polypectomy (specimen)*
___ Segmental resection *[G-8424, 397055003] Specimen from small intestine obtained by segmental resection (specimen)*
___ Whipple resection *[G-8425, 397056002] Specimen from small intestine obtained by Whipple resection (specimen)*
___ Other (specify) *not coded*
___ Not specified *[G-8360, 122638001] Tissue specimen from small intestine (specimen)*

The CAP cancer checklists standardize the format for cancer pathology reporting. Encoding the checklists with SNOMED CT standardizes the meaning of the items on the checklists. This should result in more complete, accurate and retrievable cancer data.

© 2007
College of American Pathologists
SNOMED Terminology Solutions

- **Margins (check all that apply)**
- Proximal Margin
- ___ Cannot be assessed
- ___ Uninvolved by invasive carcinoma
- ___ Involved by invasive carcinoma
- ___ Adenoma absent at proximal margin (for appendectomy specimens)
- ___ Adenoma present at proximal margin (for appendectomy specimens)

- Specify grade of dysplasia: _____

Answer 2: A computer:

- Can't distinguish the question from the answer(s). What is the question being asked here?
- Can't tell that the first selection invalidates all other selections
- Can't tell that the proximal margin's location (i.e. the definition of the "Proximal Margin") changes depending on the type of resection: Proximal appendix (en-face) for appendectomy vs the distal ilium for (hemi)colectomy. These are actually different types of margins.
- Does not know that absent / present and uninvolved / involved are paired questions. You can't have both absent and present checked!

© 2007
College of American Pathologists
SNOMED Terminology Solutions

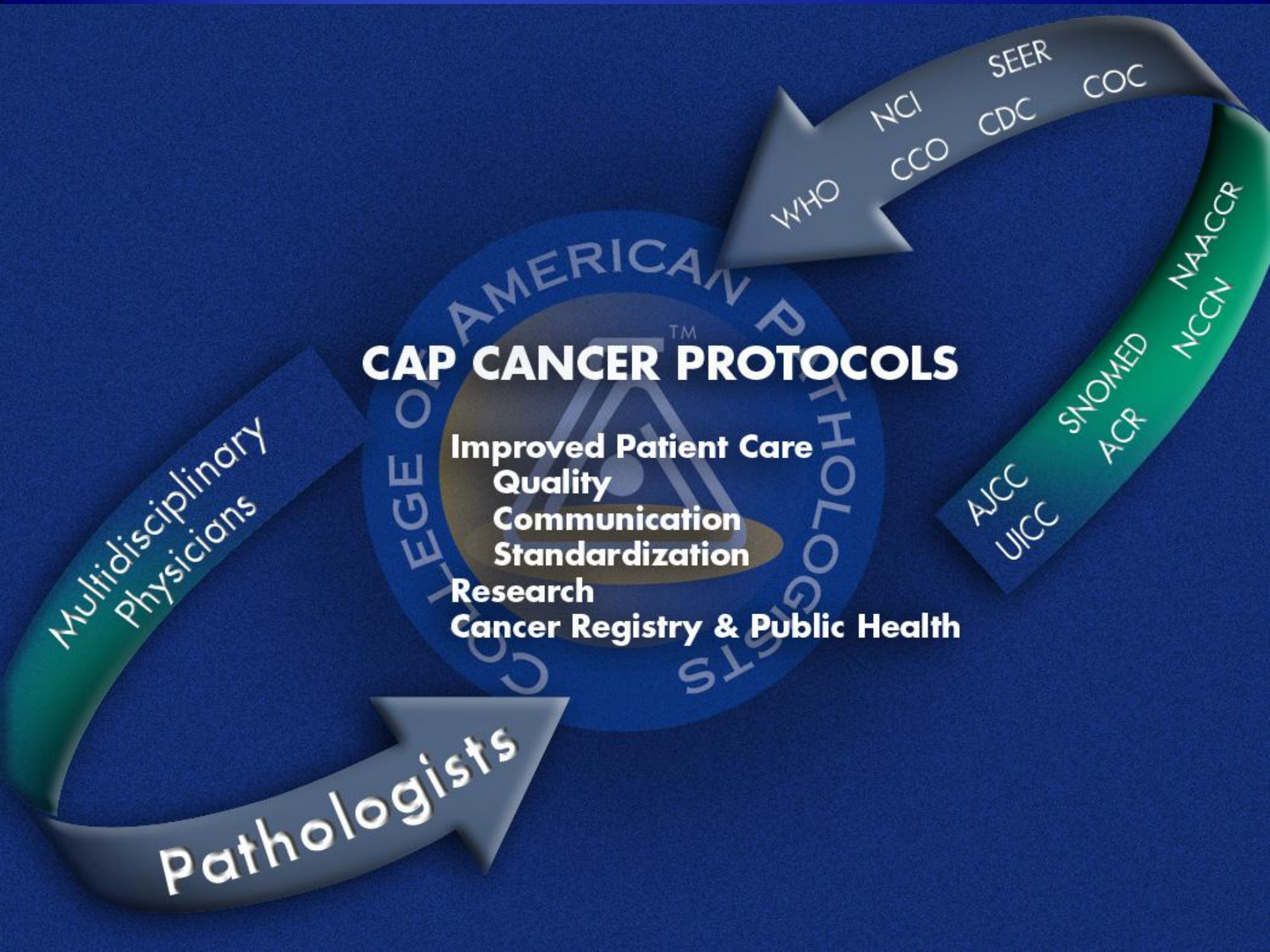
- **Margins (check all that apply)**
- Proximal Margin
- ___ Cannot be assessed
- ___ Uninvolved by invasive carcinoma
- ___ Involved by invasive carcinoma
- ___ Adenoma absent at proximal margin (for appendectomy specimens)
- ___ Adenoma present at proximal margin (for appendectomy specimens)

- **Specify grade of dysplasia**: _____

Answer 2 (cont.): A computer:

- Can't effectively process free text and still allow data mining for QA, surveillance, or research
- Has no way to enforce a logical and correct set of answers for free text
- Is subject to variant spellings and misspellings
- Has no standard way to transmit the data items in a consistent manner to central repositories (e.g. registries, etc)
- Has no way to know how to format the question/answer pairs for the end-user. What on screen "controls" will be used?
- Etc., etc.

© 2007
College of American Pathologists
SNOMED Terminology Solutions


CAP CANCER PROTOCOLS
Improved Patient Care
Quality
Communication
Standardization
Research
Cancer Registry & Public Health
Multidisciplinary Physicians
Pathologists
WHO
NCI
CCO
SEER
CDC
COC
AJCC
UICC
SNOMED
ACR
NAACCR
NCCN

The CAP Cancer Protocols – Standard of Care in Cancer Centers

- The **American College of Surgeons - Commission on Cancer** (ACS-CoC) has recognized the value of the CAP cancer checklists in caring for cancer patients.
- Beginning January 1, 2004, the ACS-CoC mandated new standards through its Cancer Center approvals program.
- One new standard requires that pathologists at ACS-CoC-approved cancer programs include all scientifically validated or regularly used data elements of the checklists in their reports for each site and specimen.
- This requirement for a "standard" in cancer pathology synoptic reporting opens the door to creating *interoperable* and *standardized* clinical documents for pathology and oncology.
- These interoperable documents can then be used by any computer system for use in patient management (e.g. exchange of electronic patient records), or aggregated and queried for research studies (e.g. caBIG, cancer registries), in addition to their use for quality assessment of cancer centers (e.g. ACoS).

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Protocol Collaboration

Histologic Type (Note B)

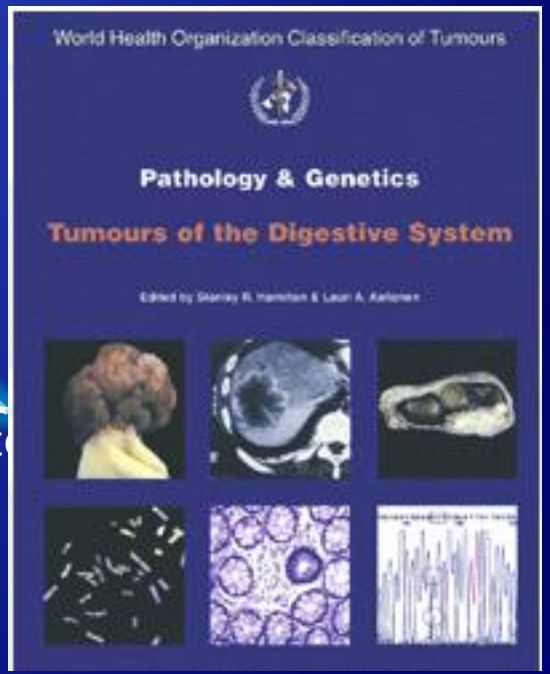
___ Adenocarcinoma
___ Mucinous adenocarcinoma
___ Signet-ring cell carcinoma
___ Small cell carcinoma
___ Squamous cell carcinoma
___ Adenosquamous carcinoma
___ Medullary carcinoma
___ Undifferentiated carcinoma
___ Other (specify): _____
___ Carcinoma, type cannot be determined

Histologic Grade (Note C)

___ Not applicable
___ Cannot be determined
___ Low-grade (well differentiated to moderately differentiat
___ High-grade (poorly differentiated to undifferentiated)

Creators, Adaptors, Adopters, Users…:

- **CAP - Cancer Committee**
- **CAP - SNOMED Terminology Services**
- **WHO (Blue Book Terminology, ICD-O3+, ICD-11)**
- SNOMED CT (IHTSDO)
- ACoS (AJCC, CS)
- ASCO
- CDC (NPCR, NAACCR)
- NCI (SEER, caBIG)
- Pathologists and Oncologists
- Others…



© 2007
College of American Pathologists
SNOMED Terminology Solutions

Computerizing the CAP Cancer Protocols

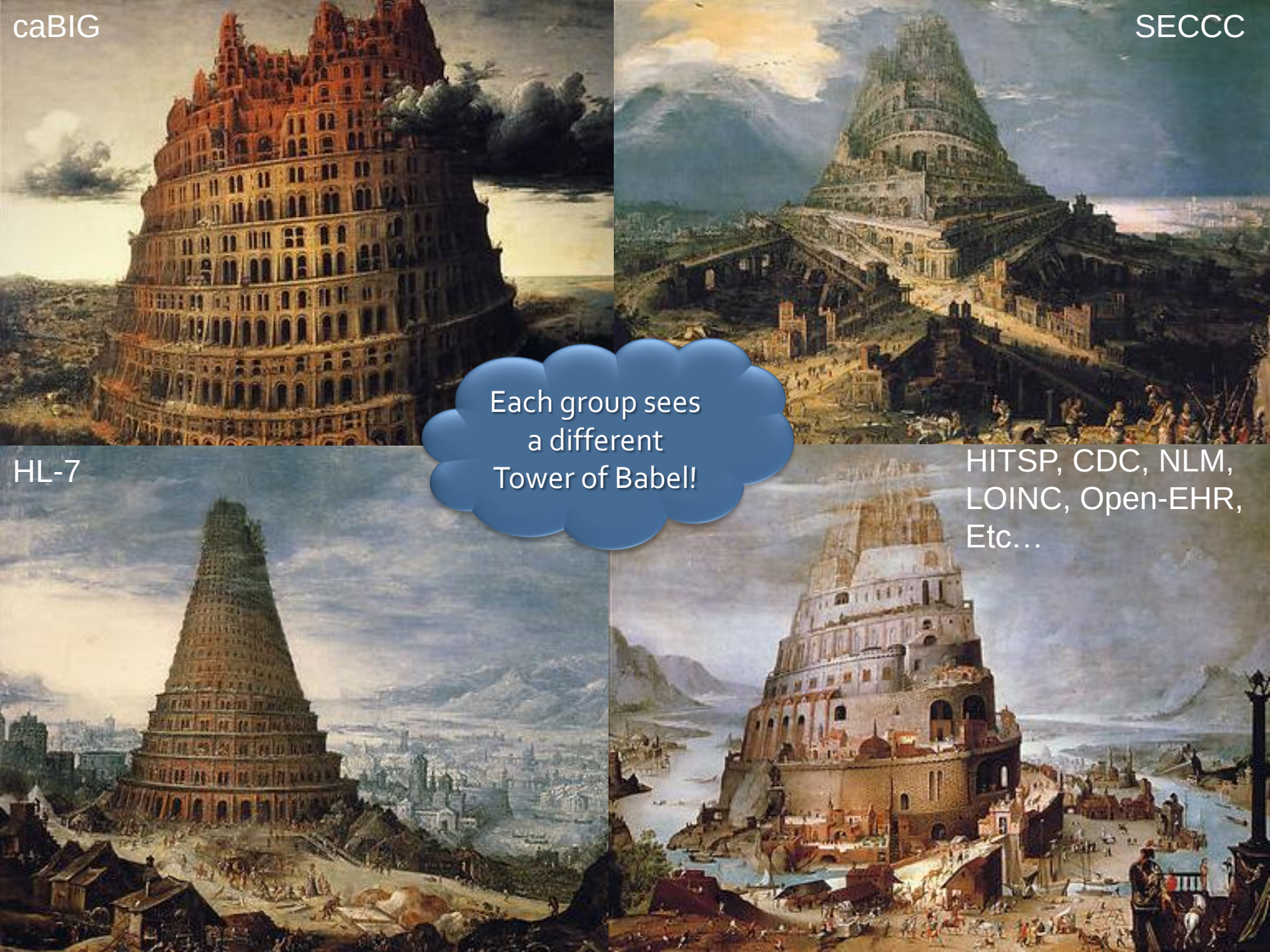
- The new *computerized* version of the Protocols, named the **SNOMED CT - Encoded CAP Cancer Checklists** (SECCC), was first released in Jan 2007 by CAP-STS
- The SECCC format is designed to address a number of problems with the paper format and to be flexible enough to accommodate rapid future change with the minimum amount of developer involvement
- This new format will be particularly important in allowing users to switch to new SECCC versions that incorporate new WHO, AJCC, CS and NAACCR data elements

© 2007
College of American Pathologists
SNOMED Terminology Solutions

SECCC - Long-Term Goals

- **GUI**: Allow multiple centers to present SECCCs to end-users (pathologists) in a consistent and interoperable manner, enabling the collection of meaningful and comparable data
- **Public Template Model:** The metadata that defines the content and presentation of the SECCC templates will be publically available on the Internet.
- **Data Transmission:** Enable multiple centers to transmit, receive, and interpret data, enabling collaborative QA, surveillance, and research efforts
- **Standard Formats:** Standardizing on SECCCs for data collection will enable groups like NAACCR to efficiently collect and analyze vast amounts of SNOMED-encoded data without the need for manual data extraction and conversion
- **caBIG Integration:** Enable distributed (federated) SECCC data repositories to be queried via caBIG
- **EMR Standardization via SNOMED:** Expansion to other SNOMED-based standard EMR forms, enabling interoperable Question/Answer Sets (QAS)

© 2007
College of American Pathologists
SNOMED Terminology Solutions


caBIG
SECCC
Each group sees a different Tower of Babel!
HL-7
HITSP, CDC, NLM, LOINC, Open-EHR, Etc…

Interoperability Levels

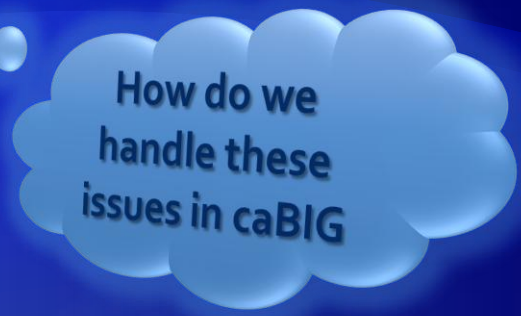
- **Syntactic**: Technical format that allows exchange, accumulation and aggregation of data and information (e.g. HL-7, XML, CDA, SOAP…)

The difficult part!

- **Semantic**: The meaning of the data elements (question and answer items) may be related to:
  - Relationships between QAS (template) versions
  - Concept definition (SNOMED CT, LOINC, etc)
  - Concept relationships (SNOMED CT concept model)
  - Concept context (encoded by template metadata):
    - Relation of Q/A items between different template versions and between different template types (e.g. Histologic Type)
    - Relation of Q/A items to base item (inheritance)
    - Injection of other templates and sub-templates after specific responses
    - "Skip areas" – item dependencies
    - Algorithms that modify data
    - Parent-child concept hierarchy with QAS
    - GUI presentation standards for QAS

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Extra Interoperability Issues: SECCC/caBIG-specific



- **Syntactic –**
  - UML Modeling
  - caDSR style
  - Template/Form mechanisms – Public repository for extra metadata
  - Global Model Exchange (GME) for data models
  - Grid enabling
  - **Where do we store all this extra metadata?**
- **Semantic**: The meaning of the data elements (question and answer items) may be related to:
  - Relationships between QAS (template) versions
  - Concept definition (conflict b/t EVS versus caDSR –DEC, VD, CDE)
  - Conflicts b/t multiple terminologies
  - Concept relationships (EVS versus caDSR –DEC, VD, CDE)
  - Concept context (encoded by template metadata):
    - Relation of Q/A items between different template versions and between different template types (e.g. Histologic Type)
    - Relation of Q/A items to base item (inheritance)
    - Injection of other templates and sub-templates after specific responses
    - “Skip areas” – item dependencies
    - Algorithms that modify data
    - Parent-child concept hierarchy with QAS
    - GUI presentation standards for QAS

© 2007
College of American Pathologists
SNOMED Terminology Solutions

QAS Project Design Goals:

- Create standard for representing the template
  - Similar to Open-EHR/Archetypes, CDA/RIM, but combining their best features with standard terminologies and enhanced functionality into a single simplified model.
- Create user-friendly TOOLS for editing and viewing the QAS templates
- Create standard and TOOLS for JIT or static conversion of the template into a data-entry form (DEF)

© 2007
College of American Pathologists
SNOMED Terminology Solutions

QAS Information Model Highlights

- Model a generic QAS Report Template, NOT a piece of tissue or patient findings, or a disease.
- QAS templates are modeled after logically-designed Q/A paper forms, but modified to support common computer paradigms (e.g. combo boxes)
- QAS Templates consist of primitive Questions, Answers, Headers, Notes, plus metadata. More primitive types are possible (e.g. tables, images)
- All answer choices must be associated with a parent question, in a Q/A hierarchy
- Each question and each answer choice may have child Q/A sets

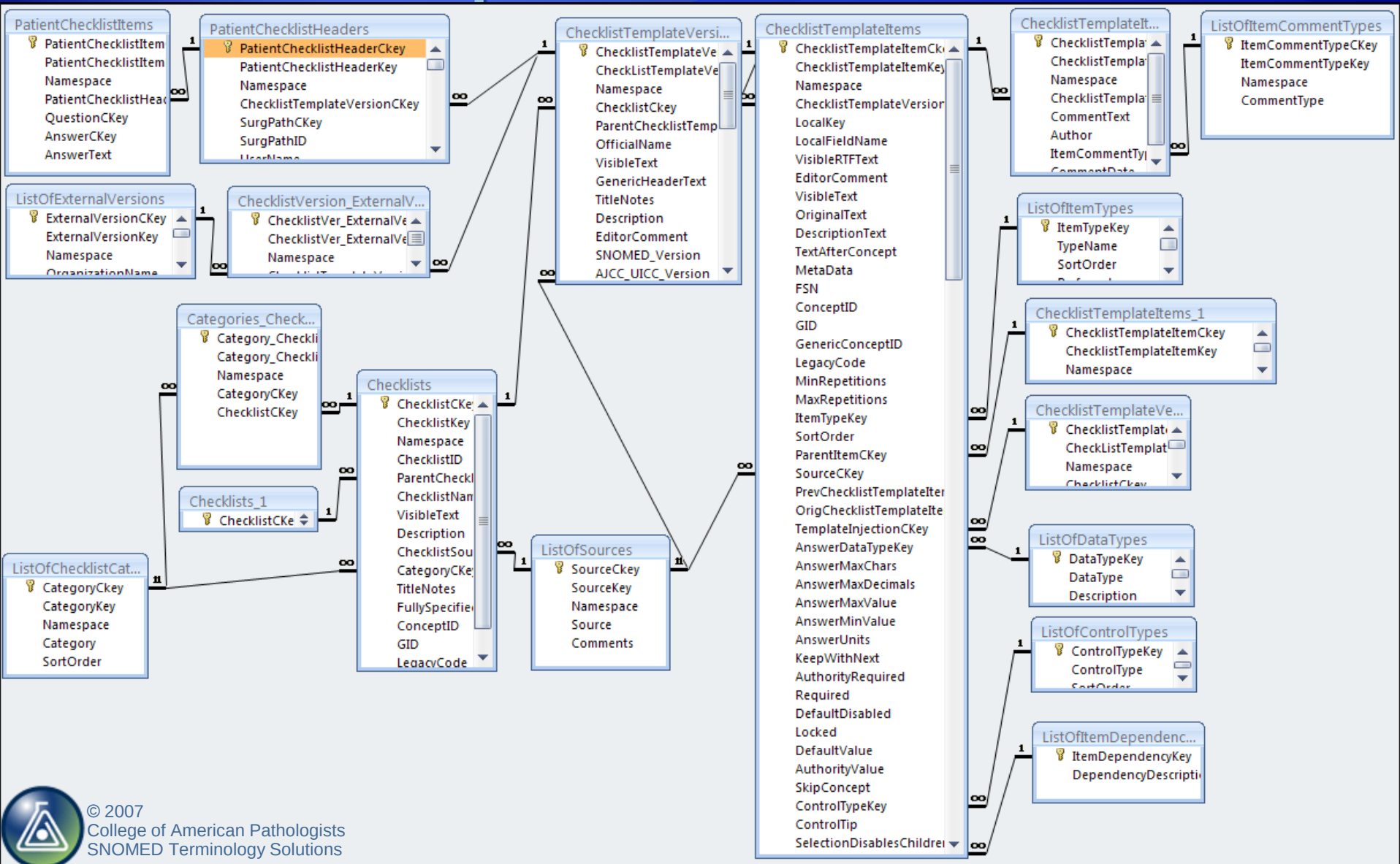
© 2007
College of American Pathologists
SNOMED Terminology Solutions

QAS Data Model Goals:

- Create SIMPLE database and XML standard for representation of QAS templates (intentionally hierarchical, rather than object-oriented or normalized)
- Create SIMPLE database and XML standard for QAS data storage
- Create SIMPLE XML standard for data transmission and interchange
- Support customized transformation into data structures/warehouses that enhance end-user querying
- Support national aggregation and querying of data through BioSense, caGRID, …

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Interoperable Templates: Template Data Model

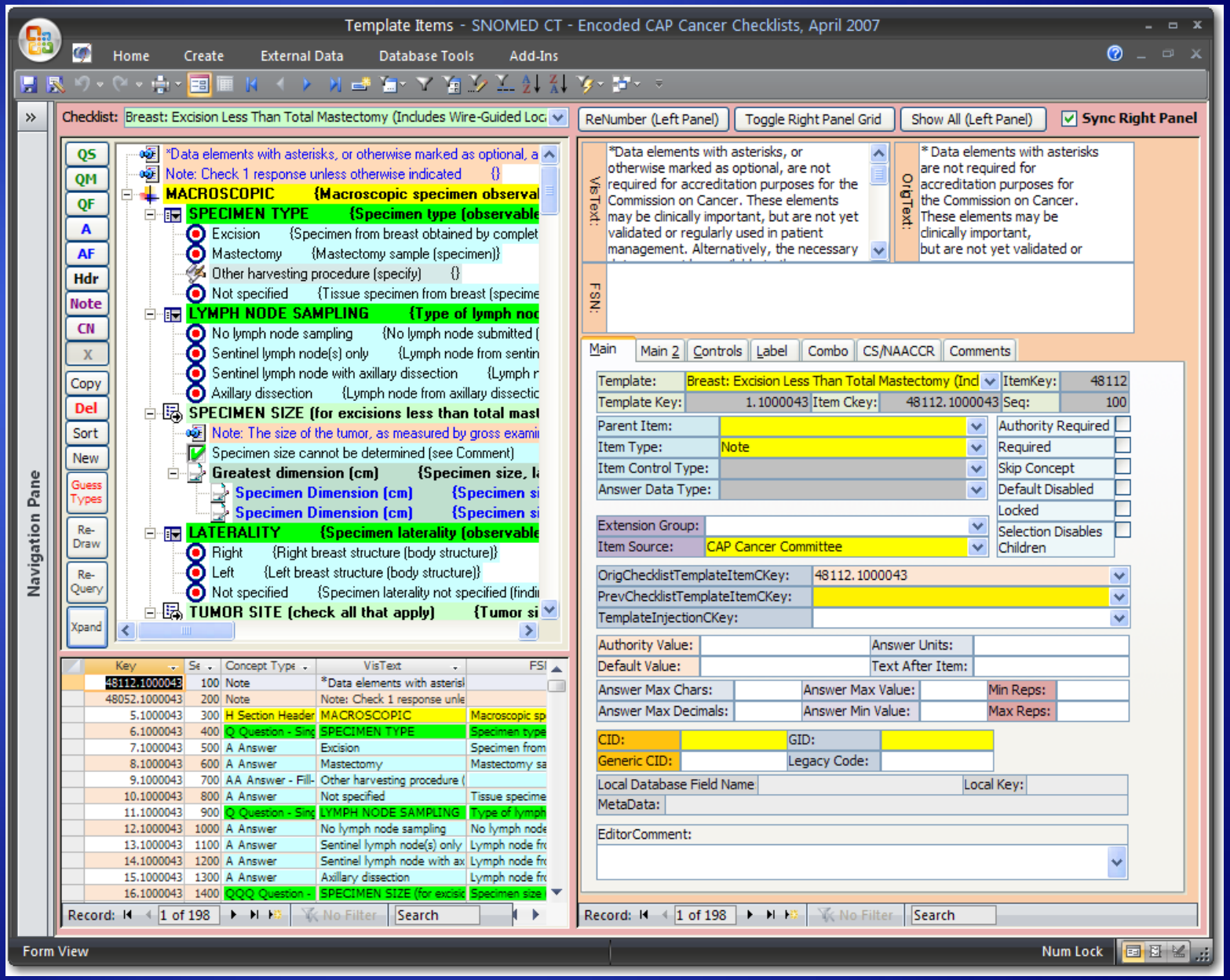


© 2007
College of American Pathologists
SNOMED Terminology Solutions

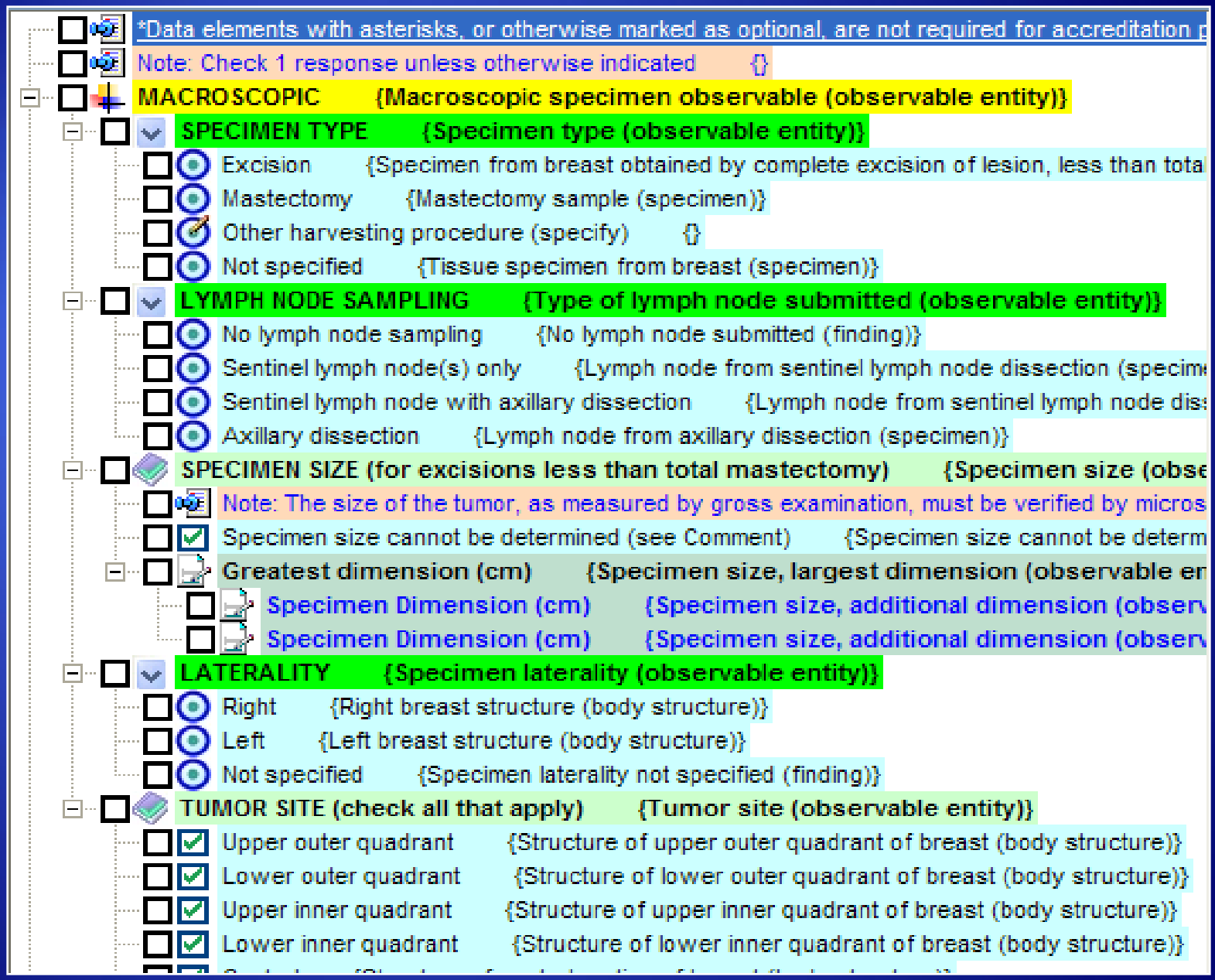
Some Advantages of the Hierarchical Model

- Self-describing during hierarchy creation
- Uses the idea of inherited "base templates," similar to OO base class, as an implementation of the caBIG proposed "Backbone Model"
- Templates or sub-templates can be injected into a QAS depending on user interaction: "object" model is fluid during QAS data entry
- Preserves all semantic context and much DEF functionality
- Allows alternate graphs through the QAS, while preserving context
- Hierarchical templates are very easy to produce, and they all share an identical UML pattern
- Template Editor tool is already available
- Supports rich metadata model that can be used for JIT DEF generation

© 2007
College of American Pathologists
SNOMED Terminology Solutions

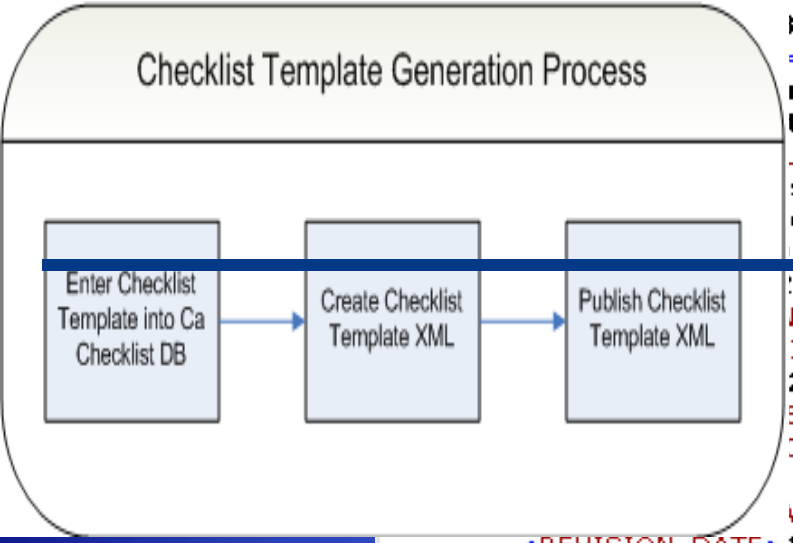


© 2007
College of American Pathologists
SNOMED Terminology Solutions



© 2007
College of American Pathologists
SNOMED Terminology Solutions

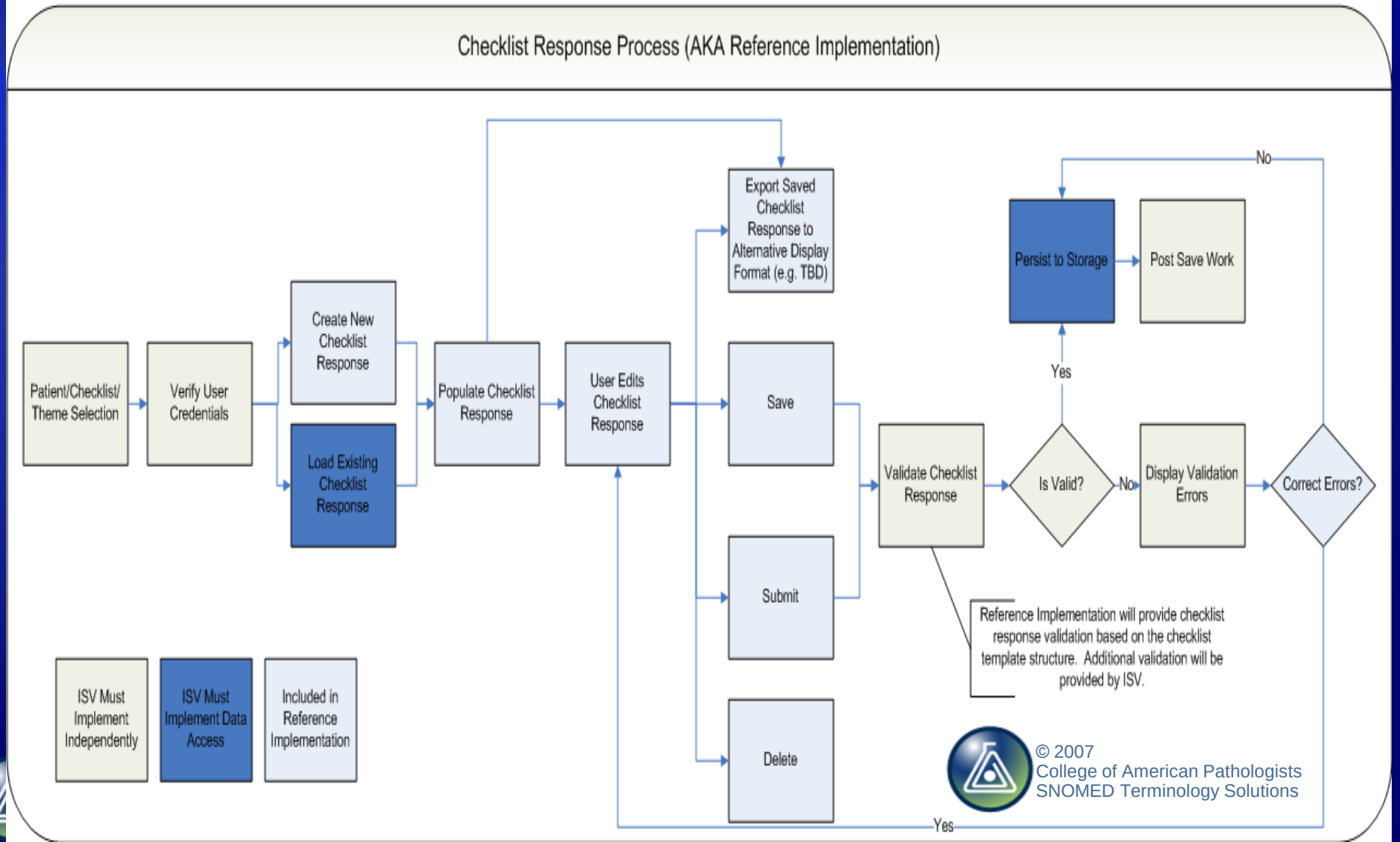
Convert template specification to computer-readable XML document for each checklist




```
<?xml version="1.0" encoding="utf-8" ?>
- <SAMPLE>
  <!-- Checklist document root sample -->
  - <CHECKLIST_TEMPLATE TEMPLATE_ID="1" VERSION_ID="1" TEMPLATE_CONCEPT_ID="406030002">

                                                    </APPLICABILITY>
                                                    </CATEGORY>
                                                    r Checklists</GROUP>
                                                    t: Protocol applies to all invasive carcinomas of the
                                                    E>
                                                    st</SHORT_TITLE>
                                                    e are notes.</TITLE_NOTES>
                                                    NS>
                                                    3</CAP_VERSION>
                                                    >January 2005</SNOMED_VERSION>
                                                    ION>123</AJCC_UICC_VERSION>
                                                    23</FIGO_VERSION>
                                                    STAGING_VERSION>123</COLLABORATIVE_STAGING_VERSION>
                                                    ONS>

                                                    TE>1/1/1900</WEB_POSTING_DATE>
        <REVISION_DATE>1/1/1899</REVISION_DATE>
        <EFFECTIVE_DATE>1/1/1900</EFFECTIVE_DATE>
        <RETIREMENT_DATE>1/1/1980</RETIREMENT_DATE>
        <APPROVAL_STATUS>Approved</APPROVAL_STATUS>

          <CAP_ALTERNATIVE_FORMAT_URL>http://www.cap.org/downloadMe.pdf</CAP_ALTERNATIVE_

          <SNOMED_ALTERNATIVE_FORMAT_URL>http://www.cap.org/downloadMe.pdf</SNOMED_ALTER
      </PUBLISHER>
    </TEMPLATE>
```

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Programming of JIT Screen Generator



Computerized Checklists in a Web Browser

Surgical Pathology Cancer Case Summary (Checklist)

College of American Pathologists Cancer Checklist; Breast: Excision Less Than Total Mastectomy (Includes Wire-Guided Localization Excisions); Total Mastectomy, Modified Radical Mastectomy, Radical Mastectomy

Protocol revision date: January 2005

Applies to invasive carcinomas only

Based on CAP, 1/31/2007

Based on SNOMED, 1/31/2007 12:00:00 AM

Breast Excision Less Than Total Mastectomy

Total Mastectomy

Modified Radical Mastectomy

Radical Mastectomy

Excision of breast tissue

MACROSCOPIC

SPECIMEN TYPE: Other harvesting procedure *Required.

LYMPH NODE SAMPLING:

SPECIMEN SIZE (for excisions less than total mastectomy)

☐ Specimen size cannot be determined (see Comment)

Note: The size of the tumor, as measured by gross examination, must be verified by microscopic examination. If there is a discrepancy between gross and microscopic tumor measurement, the microscopic measurement of the invasive component takes precedence and should be used for tumor staging.

Greatest dimension

Specimen Y Dimension

Specimen Z Dimension

LATERALITY:

TUMOR SITE (check all that apply)

☐ Upper outer quadrant

☐ Lower outer quadrant

☐ Upper inner quadrant

☐ Lower inner quadrant

☐ Central

☐ Not specified

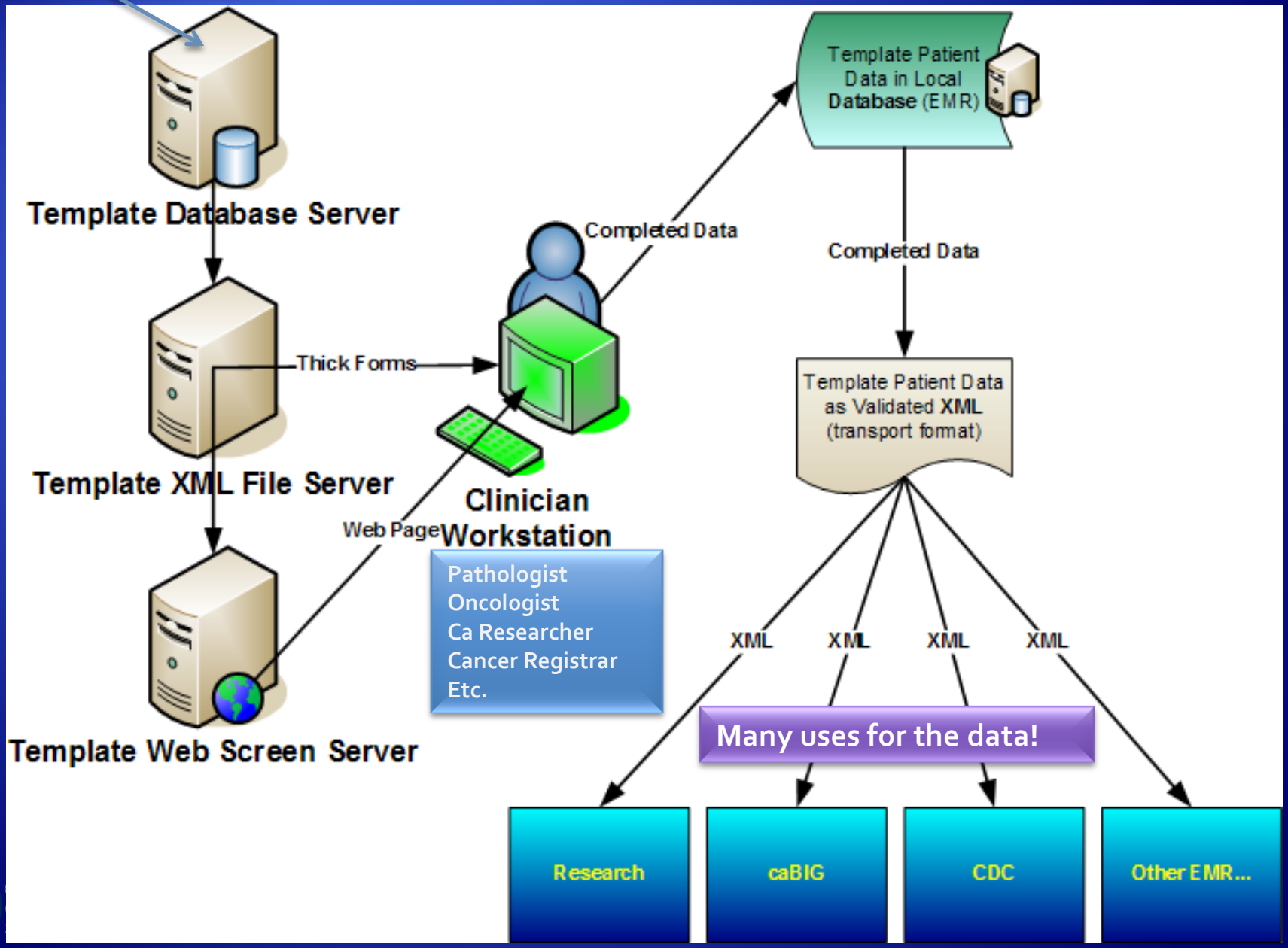
© 2007
College of American Pathologists
SNOMED Terminology Solutions

CAP Cancer Protocols – Workflow

- Collaborative creation of new and modified Protocols
  - Use and creation of new terminology as needed (WHO, IHTSDO / SNOMED)
- Paper protocols placed on CAP web site
- Conversion of Protocol to computerized template
- Creation and assignment of SNOMED CT concepts (and CS, NAACCR, LOINC and others as well)
  - SNOMED CT codes are natively linked to ICD-O3 codes, and this linkage needs to be updated with each change in WHO terminology.
- Conversion of computerized Protocol to XML distribution format
- Distribution of new Protocol to adopters (e.g. Vendors, caBIG, Cancer Centers, etc)
- Adopters process new protocols for use in local systems
- New Protocols used by pathologists, oncologists, researchers, registrars, etc.
- Patient data from completed Protocols is used to guide patient care, research, quality reviews…

© 2007
College of American Pathologists
SNOMED Terminology Solutions

The Big Picture



Needs for Interoperable QAS Templates

- EMR systems
- Patient personal health records
- Health maintenance records
- Public health reporting
- Nursing, RT, OT, PT …
- Immunization databases
- Infectious disease surveillance
- E-prescribing
- Asthma care, diabetes care, clinical trials, etc, etc
- Disaster management

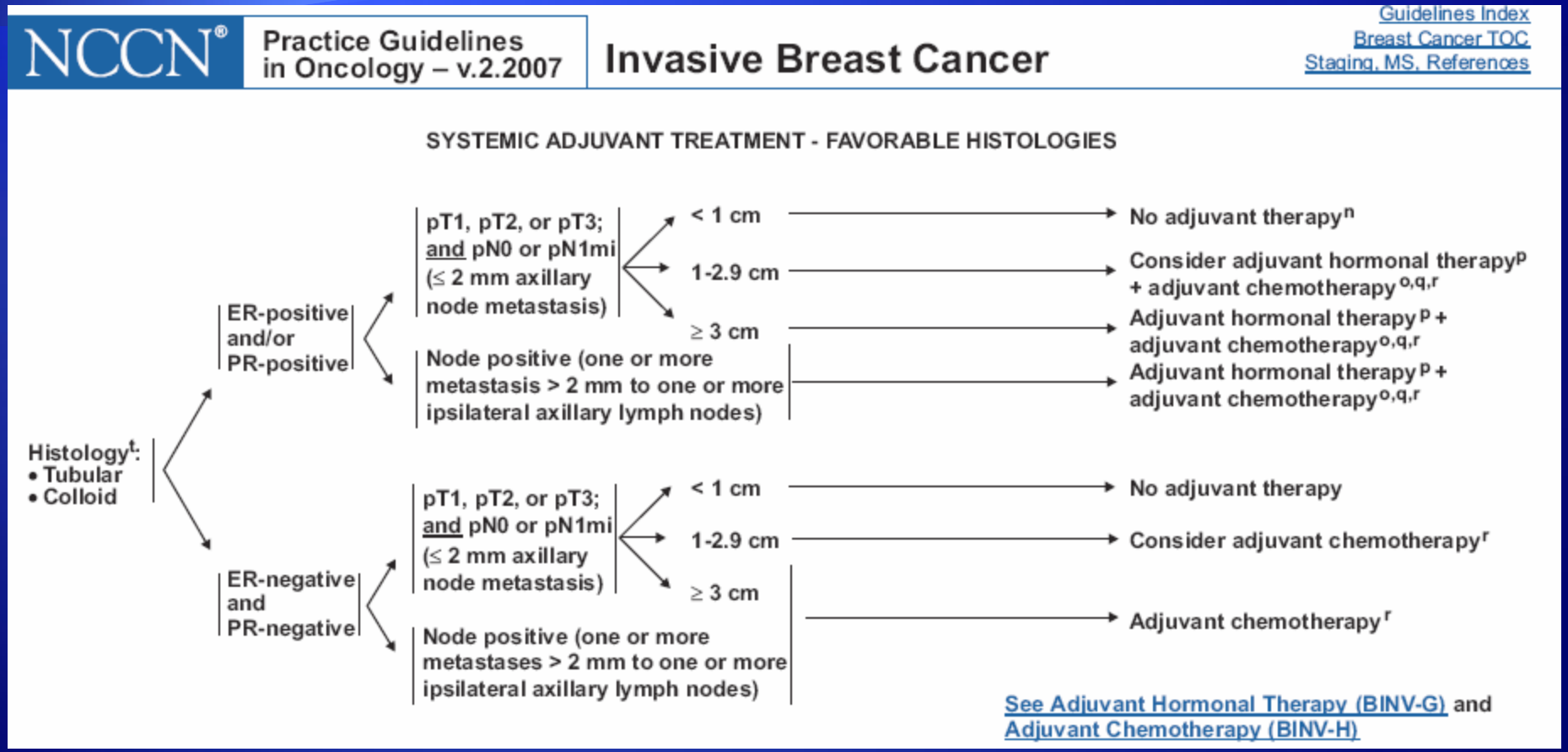
© 2007
College of American Pathologists
SNOMED Terminology Solutions

Usefulness of SECCC Templates and Patient Data

- Patient pathology data derived from SECCC templates are useful for almost every part of caBIG
- The standardized, metadata-rich data elements from the SECCC templates will be a valuable and heavily-used addition to the caDSR and the NCI Thesaurus.
- The SECCC templating model can serve as a prototype for other caBIG projects that need customizable Question/Answer Sets (QAS), and fast QAS template generation.

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Complex QAS Algorithm Support: Divergent & Convergent Pathways



© 2007
College of American Pathologists
SNOMED Terminology Solutions

Basic End-User Goals of the Proposed caBIG-SECCC Project

- Get de-ID'd patient data derived from the CAP Cancer Protocols on the Grid
- Directly annotate other caBIG project data sets with SECCC data
- Allow grid-enabled data to flow seamlessly to research groups such as NAACCR/NPCR and SEER
- ?Create DEF JIT generator than runs off of the DSR (UML Model)?
- Allow robust querying of grid data, joined with other caBIG data sets (e.g. tissue banks, clinical trials, microarrays…)
  - Link via honest broker identifier?
- Create end-user software tools to enhance grid enabling of remote sites and easy grid querying of SECCC-derived data elements

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Why the SECCC project is different

- Requires frequent changes (new templates, new versions, customization) to the template QAS (this is NOT necessarily the same as the information or UML model)
- Requires a robust template versioning system that allows querying through multiple and selectable versions
- Every template follows the same fixed, but highly flexible information model, via modified Model-View-Controller design patterns

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Why the SECCC project is different

- Templates encode the metadata required to generate a complete *functional* data-entry form (DEF). Standard presentation is critical to contextual semantics and interoperability.
- May require that hundreds of new CDEs be added to the DSR, and ideally also the NCI Thesaurus. (Who will curate this???)
- Significant amounts of metadata (e.g. for presentation) need not be referenced directly in the UML model or "form builder" model; Instead it may be encoded in external XML files or XML blobs within a form model.
- The scope of the project precludes manual creation of UML models and forms for each template.
- Requires coordination with ISVs and home-grown systems.
- Requires advanced and simple tools to aid widespread integration.

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Stakeholders

- College of American Pathologists, caBIG
- Cancer Surveillance:
  - CDC, NPCR, NAACCR, NCI/SEER, Cancer Care Ontario, many others
- WHO – Blue Books/Tumor classifications, ICD-O3+, ICD-11
- American College of Surgeons
  - AJCC
  - Collaborative Staging
- Pathology Informatics Vendors
- Pathologists
- Oncologists
- Cancer Researchers
- Path Informatics Systems and EMR Vendors
- Standards Development Organizations
- PATIENTS!

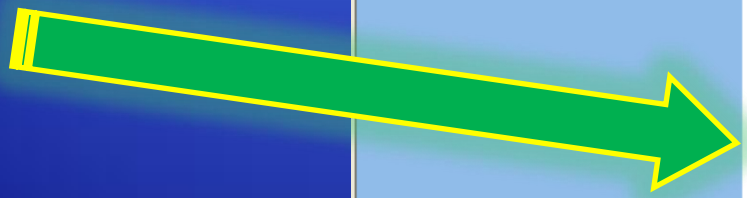
© 2007
College of American Pathologists
SNOMED Terminology Solutions

Licensing of the CAP Cancer Protocols

© 2007
College of American Pathologists
SNOMED Terminology Solutions

CAP Cancer Protocols:

Terms of Use



© 2007
College of American Pathologists
SNOMED Terminology Solutions

User ID:

Password:

Remember me on this computer

Login

Hint

Forgot your user ID or password?

Need help logging in?

Create an Account

Cancer Protocols and Checklists

Updated September 17, 2007

The College of American Pathologists (CAP) publishes and owns the copyright in the CAP Cancer Protocols (the Protocols). The CAP hereby authorizes use of exact copies of the Protocols by physicians and other health care practitioners in reporting on surgical specimens for individual patients, in teaching, and in carrying out medical research for non-profit purposes.

The CAP also authorizes physicians and other health care practitioners to make modified versions of the Protocols solely for their individual use in reporting on surgical specimens for individual patients, teaching, and carrying out medical research for non-profit purposes.

The CAP further authorizes the following uses by physicians and other health care practitioners, in reporting on surgical specimens for individual patients, in teaching, and in carrying out medical research for non-profit purposes:

(1) **Dictation** from the original or modified protocols for the purposes of creating a text-based patient record on paper, or in a word processing document.
(2) **Copying** from the original or modified protocols into a text-based patient record on paper, or in a word processing document.
(3) The use of a **computerized system** for items (1) and (2), provided that the Protocol data is stored intact as a single text-based document, and is not stored as multiple discrete data fields.

Other than uses (1), (2), and (3) above, the CAP does not authorize any use of the Protocols in electronic medical records systems, pathology informatics systems, cancer registry computer systems, computerized databases, mappings between coding works, or any computerized system without a written license from CAP. Applications for such a license should be addressed to the SNOMED Terminology Solutions division of the CAP.

Any public dissemination of the original or modified Protocols is prohibited without a written license from the CAP.

caBIG Use of the CAP Cancer Protocols

- The College of American Pathologists (CAP) owns the copyright on the CAP Cancer Protocols. The Protocols are researched and written by the CAP Cancer Committee.
- Licensing of the Protocols is done through CAP - SNOMED Terminology Solutions (CAP-STS).
- Licensing of SNOMED CT, which is used in the encoded version of the Protocols, is available through an IHTSDO Affiliate License. This is free in the US (from the NLM) and in other IHTSDO member countries.
- Organizations, developers, and end-users wishing to implement the Protocols in a computerized system must receive written authorization from CAP-STS.

© 2007
College of American Pathologists
SNOMED Terminology Solutions

caBIG Usage of the CAP Cancer Protocols

- caBIG and its participants who work with the Protocols must work with CAP-STS to ensure:
  - The content of the caBIG-implemented Protocols accurately represents the intent of the expert authors (The CAP Cancer Committee) and CAP's **P**athology **E**lectronic **R**eporting **T**askforce (PERT)
  - A single standard for robust interoperability
  - Implementation of the Protocols in a standard manner appropriate for use by:
    - ACoS cancer center certification
    - Cancer registry organizations (NAACCR, CDC/NPCR, NCI/SEER, CCO, etc.)
    - The new AJCC 7th edition TNM staging system
    - Collaborative Staging (CS), as used by NCI/SEER
    - Automated staging algorithms as implemented by the CDC, in concert with CS
  - Protection of
    - CAP's copyright for the Protocols
    - CAP's exclusive right to charge fees for Protocol licensing and consulting

© 2007
College of American Pathologists
SNOMED Terminology Solutions

caBIG Use of the CAP Cancer Protocols

- Subject to the above conditions, CAP-STS will license Protocol use to all caBIG participants, at no charge, for official caBIG-supported development and testing purposes only.
  - Each user wishing to use the Protocols for caBIG development or testing must obtain a free institutional or personal license for this purpose from CAP-STS. Email SECCC@cap.org
- Use of the Protocols for actual computerized data storage, patient management, or research purposes will require an additional license from CAP-STS. CAP-STS will work with licensees to provide reasonable licensing terms.

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Breakout Group 1: Vocab/Architecture Gurus

Modeling the Protocol templates in UML/caDSR/Form Tool/EVS/SNOMED CT

- Can the SECCC template metadata be represented in the current system? What modifications need to be made?
- How will the UML represent the hierarchical and inheritable/injectable template structure of the protocols?
  - Can the UML modeling step be bypassed, with direct creation of caDSR/EVS objects?
  - Can/Should the UML model be created via a direct XMI transform from the hierarchical template model?
  - Are there methods of programmatically creating DECs, Value Sets, CDEs from the template metadata? Is it necessary to got through UML for this first?
  - Are there available methods to directly import the terminology (SNOMED CT), and create EVS entries?
  - Can SNOMED be used as a native vocabulary, like NCI-T?
- What new standard terms do we need in the DSR to support the SECCC, e.g. Representation terms, categories, classifications, template versions, etc.

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Breakout Group 2: Clinical Use-Case Gurus

Developing a Framework for Clinical Datasets Beyond the Pathology Synoptic Reports (CAP Protocols)

- Many essential clinical oncology fields are not represented as minimal data sets or synoptic reports. We need specific and standardized synoptic questions with defined standard answer choices for each clinical area
  - Surgical oncology, Med/Peds/Gyn Oncology, Tumor Registry (NAACCR), Collaborative Staging, Molecular diagnostics, Chemo protocols, related treatments, stem cell transplant (NMDP/ABMTR) forms, radiology, RadOnc, etc. Task: Add more groups and specific question sets as time permits. Prioritize the list.
  - Who will form the groups to produce these sets? Who will review them? Who will implement them in caBIG? Inter-Organizational structure?
  - Can these Q/A sets adopt the CAP Cancer Protocol template format, or is there a need for other formats? Are there other use cases that would require modifications to the SECCC template model?
  - What types of complex data are required for synoptic clinical annotation: do we really need x-rays, microarray data sets, path images, or is a summary adequate for clinical annotation (at this stage)? Is there a standard for radiology synoptic reporting? What standards are available for outcomes reporting and therapy selection in Oncology? Severity of Disease reporting? Adverse Reaction Scoring Systems? Cardiac Performance Status? Renal Status? Exercise Tolerance? Others?
  - Some other caBIG groups may be producing clinical annotations that could be "harvested" as synoptic Q/A pairs (e.g. microarray, radiology). Where do these exist, and how can these be converted to a standard Q/A structure appropriate for general clinical annotation and general synoptic reporting? Is there any caBIG organizational structure to bring this data into CAE etc? Can we implement a standard format for all forms of clinical annotation, similar to the SECCC?
  - In creating UML/DEC models for the Protocols and similar "checklists," should we use a flat model (No complex objects, just properties of a generic CAP_Protocol object class), or alternatively, should we create object classes (e.g. CAPTumor, CAPTumorCells, CAPTumorMargins, TNM_Status, CS_Extent, etc.). In other words, it it sufficient to leave the object structure to the terminology model (NCIt or SNOMED CT) for representation?
  - What header information (UML/CDEs) is needed to identify the subject of each Protocol's data set?

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Breakout Group 3: Question/Answer Sets (QAS) in Practice: Building the Tools

Critical need for user-friendly, fast tools: How can we organize to do this? CDC/PHIN Coordination? Priorities for these:? Other tools needed?

- **Distributed template editing**
  - Beyond the Forms tool – need to support more robust metadata-rich forms
  - Tools for creating, editing templates
  - Is there a pressing need to modify existing SECCC template structure at this point?
  - Must support hierarchical base templates, template inheritance, template injection, sub-template injection, algorithms, calculations
  - Dev language (.NET vs Java vs Access) to maximize contributors to project
  - How to connect .NET apps to existing Java APIs? caCore should support .NET or .NET must use a Java Bridge?
- **JIT Screen generator**
  - Based on above templates, need screen generator (web/thick forms/smart client) that annotates or plugs into other caBIG apps
  - Recommend an architecture and strategy based on the template model
    - Template source is web service, API, XML on public server??
    - App is web based, thick forms, smart client (click-once deployment) app??
    - Technology is roll-your-own (C#, Java)/AJAX, XAML, X-forms, PDF forms, InfoPath, ???
    - Data export formats, and direct connection to data repositories: formats?
- **Template repository**
  - How and Where will we store template versions: caDSR, Public XML, somewhere else??
- **Query generator tool**
  - Select template(s) and versions desired, and then query with generic or specific questions, using base-level CDEs when appropriate
  - Suggest architecture, SQL generation mechanisms, working with standard EAV-style data models through the Grid.

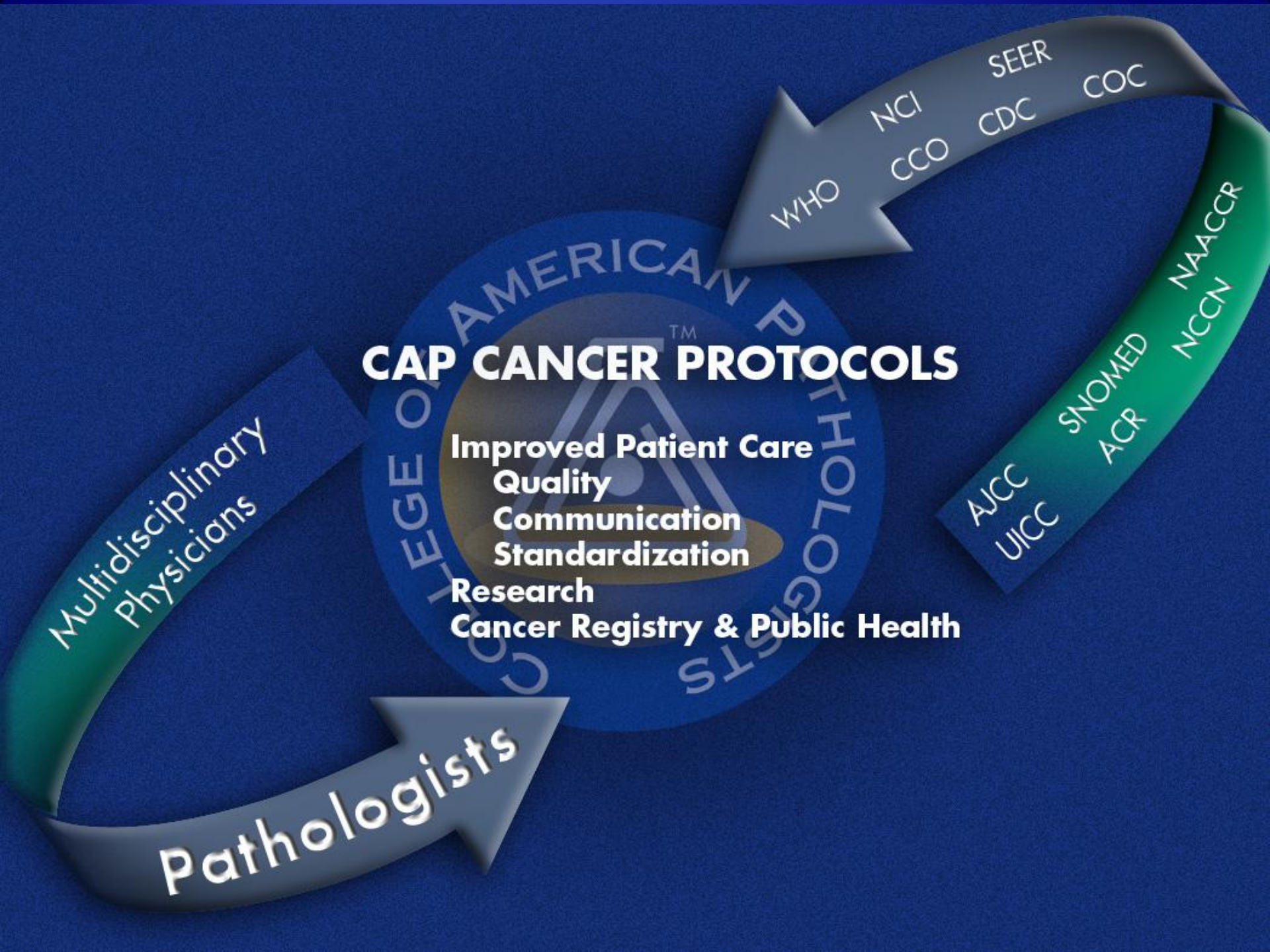
© 2007
College of American Pathologists
SNOMED Terminology Solutions

End of Introduction

© 2007
College of American Pathologists
SNOMED Terminology Solutions

More Detail for Working Group

© 2007
College of American Pathologists
SNOMED Terminology Solutions


CAP CANCER PROTOCOLS
Improved Patient Care
Quality
Communication
Standardization
Research
Cancer Registry & Public Health
Multidisciplinary Physicians
Pathologists
WHO
NCI
CCO
SEER
CDC
COC
AJCC
UICC
SNOMED
ACR
NAACCR
NCCN

Introduction to the new "Common Template Objects"

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Generic Question Objects: Samples

Cancer Pathology Template

Formatting Guidelines for CAP Cancer Protocols

DRAFT!

Guide to symbols in this document:

{…} = variable text that changes for each Protocol
[] = A check box answer choice
[] LIST = A list of specific multi-select check box choices that varies for each Protocol
> = A single-select answer choice (e.g. a combo box answer or option button)
> LIST = A list of specific single-select (e.g. combo box) choices that varies for each Protocol
! = A question that is answered with fill-in text

All questions and section headers are in bold

In some cases, single-select questions will request a text fill-in to provide details for that answer choice. These are generally marked with terms like "(specify)," "(describe)," etc.

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Changes to Section Headings

- The previously used headings "Macroscopic" and "Microscopic" are not to appear in the protocols (Committee consensus, July 2007). Try to maintain the structure and wording found in the following question/answer sets. However, not all of the headings below will appear in each checklist and some checklists may require additional headings or modified question/answer sets.
  - Indicates potential change to DEC object classes

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Generic Specimen Objects

- **Specimen(s)** (check all that apply)
  [] LIST of organs/body sites/sub-sites
  [] Other (specify)
  [] Not specified
- **Procedure** (check all that apply)
  [] LIST of procedures used to obtain specimen
  [] Other (specify)
  [] Not specified
- **Specimen Integrity** (choose 1)
  > Intact
  > Ruptured
  > Fragmented
  > Other disruption (describe)
  > Indeterminate
- **Specimen Transport Medium** (choose 1)
  > None
  > Saline
  > Formalin
  > B5
  > Other (specify)

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Generic Specimen Objects:

- **!Specimen age before processing (minutes)**
- **Specimen Handling** (check all that apply) (e.g. CNS)
  [] Squash/smear/touch preparation
  [] Frozen section
  [] Tissue for electron microscopy
  [] Frozen tissue
  [] Unfrozen for routine permanent paraffin sections
  [] Prepare for tissue banking (specify details)
  [] Other (specify)
  [] Not specified
- **Specimen Size**
  [] Size cannot be determined
  .....**!Greatest dimension (cm)**
  .....**!Additional dimension (cm)**
  .....**!Additional dimension (cm)**
- **!Specimen Weight (g)**
- **Specimen Laterality** (choose 1)
  > Left
  > Right
  > Midline
  > Bilateral
  > Multiple sites (specify)
  > Other (specify)
  > Information not available

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Generic Tumor Objects

- **Tumor Laterality** (choose 1)
  > Left
  > Right
  > Midline
  > Bilateral (single tumor)
  > Midline and Bilateral (single tumor)
  > Multiple sites, unilateral (specify sites)
  > Multiple sites, crosses midline (specify sites)
  > Multiple sites, laterality unknown (specify sites)
  > Other (specify)
  > Information not available
- **Tumor Site** (choose 1)
  > LIST of specific locations
  > Other (specify)
  > Tumor site cannot be determined
  > Not specified
- **Tumor Size**
  [] Size cannot be determined
  .....**!Greatest dimension (cm)**
  .....**!Additional dimension (cm)**
  .....**!Additional dimension (cm)**
- **Tumor Weight (g)**

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Generic Tumor Objects

- **Macroscopic Tumor Perforation** (choose 1)
  > Present
  > Absent
  > Cannot be assessed
- **Macroscopic Extent of Tumor** (e.g. Kidney, Wilms) (check all that apply)
  [] LIST of Sites
  [] Other (specify)
  [] Cannot be determined
- **Tumor Focality** (choose 1)
  > Unifocal (specify location)
  > Multifocal (specify locations)
  > Separate tumor nodules in same lobe
  > Separate tumor nodules in different lobes (specify sites): _
  > Synchronous {carcinomas} (specify sites)
  > Cannot be determined
- **Tumor Configuration** (check all that apply)
  [] Exophytic (polypoid)
  [] Infiltrative
  [] Ulcerating
  [] Other (specify)
- **Tumor Border Configuration** (may be combined with previous **Tumor Configuration**) (check all that apply)
  [] Pushing
  [] Infiltrating

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Margins (choose 1)
- Cannot be assessed
- ALL margins **UNinvolved** by {invasive carcinoma}
  - **Specify examined margins that are uninvolved by {invasive carcinoma}** (check all that apply)
    - ☐ LIST of margins
    - ☐ Other margin (specify _____)
  - **Specify closest UNinvolved margin** (choose 1)
    - LIST of margins
    - Other margin (specify _____)
  - **Distance of {invasive carcinoma} from closest margin: _____ (mm)**

  The following sub-question may sometimes need to be answered
  even when there is margin involvement (e.g. breast/DCIS)
  - **Margin involvement by {adenoma/intramucosal carcinoma}** (choose 1)
    - Margin(s) are involved by {adenoma/intramucosal carcinoma}
      - **Specify margins involved by {adenoma/intramucosal carcinoma}** (check all that apply)
        - ☐ LIST of margins
        - ☐ Other margin (specify _____)
    - ALL margin(s) are UNinvolved by {adenoma/intramucosal carcinoma}
    - Cannot be assessed

- Margin(s) **involved** by {invasive carcinoma}
  - **Margin 1** (e.g. proximal) (choose 1)
    - Cannot be assessed
    - Margin **involved** by {invasive carcinoma}
    - Margin **UNinvolved** by {invasive carcinoma}

…

  - **Margin 10** (e.g. distal) (choose 1)
    - Cannot be assessed
    - Margin **involved** by {invasive carcinoma}
    - Margin **UNinvolved** by {invasive carcinoma}

© 2007
College of American Pathologists
SNOMED Terminology Solutions

More Common Template Objects: "Base Classes"

- Much more than shown here…
- Many changes coming in:
  - TNM Staging
  - Collaborative Staging – calculated staging
  - WHO Tumor Classification
  - Molecular Markers
  - Flow Cytometry
  - Prior Therapy
  - Etc.
- Expansion of SECCC model to Surgical Oncologists, Medical/Peds/Gyn Oncologists, etc

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Fundamental Working Group Issues

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Approaches to Adopting External Projects: mzXML, BioPAX, and MAGE Examples from VCDE/Arch. Whitepaper

- Three existing projects sought to integrate with caBIG
- Compromises were required by both caBIG and the individual projects
- caBIG architecture, best practices, and tools do not fully support many use cases
- Custom solutions required for implementation in caBIG – potentially difficult and expensive

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Notable Quotes from caBIG Documents

- "Semantic interoperability cannot be sustained by normalized vocabulary and data elements alone – information model alignment is also required"
  - Comment: a third component is also needed - context
- "Semantic Interoperability: The mechanisms by which caBIG models share common object definitions"
- "Information modeling is not a trivial process"
- mzXML: "The approach taken is to both create an object model that *mimics the data model* and is semantically well-defined."
  - Comment: Our task with the SECCC is to mimic the functionality and semantics approach of the template model
  - Comment: For templates, there may be 2 data models: the template data model and the patient data model.
- BioPAX approach: "model a subset of the data in UML, and then model the entire data object as a single [XML string] object."
- "It still takes considerable effort to adopt complex, large standards. The potential mismatch between the external standard and the caBIG harmonized object model should also be taken into account when adopting the external standard, as mapping between…[them]…may require support for complex translation tools and services."

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Fundamental Issues 1

- Can SECCC templates be completely represented in UML? Do they need to be completely represented?
- Do we need to add ALL template metadata to the DSR?
  - Which metadata subset will enable the best querying experience?
  - E.g. do we want to support DEF-generation metadata in the DSR?
  - Where will we store the remaining metadata?
- Since templates may assume a complex hierarchical structure, are the Form Builder and grid-querying tools up to the task?
- Do we need to store all of the template metadata in the DSR as UML/XMI, or can we link to an XML repository on a public server?
  - Should we store an XML string metadata blob within a UML object?
  - Should we enhance the Form Tool to support SECCC-style templates?
  - Should we build a new Template Editor for this purpose?

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Fundamental Issues 2

- Templates do not specify a data storage format – where is this format to be specified (if anywhere)? In the GME? The data format, which does not match the template format, must be optimized for grid-based queries, e.g. by including versioning data for multi-version template searches.
- We should not have each site create it's own data storage model, because this means each site will then have to write custom, non-sharable query tools in order to be grid-enabled.
  - Who will build these tools?
- What is the standard data transmission format for an entire template data set?
  - Is this to be specified somewhere?
  - I suggest simple EAV (QuestionKey/AnswerKey/Fill-in) format, plus a template link.
  - The transport mechanism (e.g. XML, HL-7, CDA, …) needs to be standardized.

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Fundamental Issues 3

- The CAP Templates are only a part of the Clinical Annotation picture.
  - Who will create standardized templates for other clinical areas?
  - Are there standardization organizations that will review and approve them?
  - Who will do the work of building the vocab, CDEs curation, and template structures?

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Mimic SECCC Template Structure in caDSR and/or Forms?

Some Random Thoughts:

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Stuffing XML Blobs into the Model? Referencing Outside Public XML Sources?

- If we create a UML Model for every checklist template version, this implies that code APIs that depend on the model (e.g. in the current TB model) will become brittle and need to be recompiled whenever there is a template update or a new template.
- We need a DEF API and a querying system that are independent of template additions or modifications. These systems must be able to read the metadata from the DSR (or public XML file), and JIT-create DEFs, DEF logic rules, template data structures, or queries as needed.

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Template and UML Models

1. Create global mapping table between template metadata elements and caBIG UML requirements
2. Create global UML outline model for all templates
3. Write program to read each template, and export as XMI file
4. Write program to import XMI files into DSR

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Global UML Template Structure

- Main Objects
  - Checklists 1→oo
  - ChecklistTemplateVersions 1→oo
  - ChecklistTemplateItems
    - This hierarchical table/object holds a large amount of complexity that would ordinarily generate a different UML model for each checklist template. However, when we leave it in the hierarchical (self-referential) format, every checklist has the same information and data model. It also avoids the use of UML object inheritance (e.g. ItemBase ←Question ←SingleSelectQuestion ←ComboBoxQuestion ←SpecificQuestion), which has a number of known bugs and can get very complex.
    - Rather than break out questions, answers, notes, headers, etc as separate objects, we can treat all of them according to an "item base class," which is simply a "line item" or row in a QAS.
    - Need multiple terminology references (SNOMED, NCI) for EVERY answer choice in DEC and Value Domain!
- Lookup tables
  - ListOfDataTypes
  - ListOfSources
  - Etc.

© 2007
College of American Pathologists
SNOMED Terminology Solutions

caDSR and Hierarchical "Object Representation" - ???

- Apparently may not be compatible with conventional DSR model, where the DSR specifies the allowed Value Set for each CDE. In our use cases, the Value Set for a Question CDE (e.g. Tumor Histology) will vary markedly between templates.
- The caDSR as currently implemented seems to be incompatible with this common real-world QAS scenario. ??? In fact, it may also be incompatible with the **ISO/IEC 11179-3** Data Element Concept and Data Element Representation. This needs clarification.
- The hierarchical structure itself supplies the Values (Answer Choices) for each Question, and bypasses this limitation of the DSR and **ISO/IEC 11179**. ???
- What effect will this have on query generation for the grid???

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Storage of Patient Data

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Proposal for caBIG-SECCC Generic Data Storage Structure

- Do we need this structure as UML in the DSR? How would it be used?
- We DO need a standard data storage mechanism, so we can build generic grid querying tools that don't rely on a custom data model. This is especially true when the available QAS questions change according to user responses.
- Enhanced EAV storage format (simple data model):
  - Header 1→oo
    - Header UID
    - Template Version identifiers
    - De ID's patient identifier
    - Other IDs – Surg Path ID, etc
    - Optional date/time stamps
    - Optional validation status
    - Etc.
  - Body (repeated Q-A pairs)
    - Row Identifier
    - Header UID
    - Sort Order
    - Question Ckey or CDE Key
    - Answer Ckey or CDE Key
    - Answer Fill-in (validated string)
    - Optional fields:
      - Blob (pictures, binaries)
      - Question OriginalCKey (for cross-version searches): could be a CDE Key also
      - Answer OriginalCKey (for cross-version searches): could be a CDE Key also
      - For quicker queries: add denormalized SNOMED Concept IDs, LOINC, ICD-O3, CS key, NAACCR, etc
      - Date/time stamps

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Generic Data Storage Structure Plusses and Minuses

- + Every data model has the same structure, making it easy to create generic queries for different templates
- + Eases the creation of generic query tools
- + EAV pattern allows the inclusion of critical metadata with each row (e.g. version keys and SNOMED codes), making cross-template and cross-version querying much easier.
- + EAV data storage formats are pervasive in EMR systems, and there is a substantial literature on them.
- + Retrieved datasets in EAV format can be JIT transformed into field-modeled tables or ETL'ed into data warehouses for more efficient more detailed querying.
  - However, these derived data models may be much harder to query, if they are queried over the grid (requiring table joins or unions for each template version)
- - Querying against an EAV data model is inefficient compared to a column-per-field model (about half as fast). It is unclear how well this will perform on the grid. It is possible that it won't make a difference, if the grid itself is the limiting efficiency factor.
- - Querying against an EAV model may require numerous slow DTS lookups, although this could be done once (recreate the entire template in memory at once) at the beginning of the query procedure.
- - Creation of SQL queries against an EAV model is substantially harder than field-modeled tables. It would require a dedicated query tool.

© 2007
College of American Pathologists
SNOMED Terminology Solutions

General Issues to Consider for this Working Session

© 2007
College of American Pathologists
SNOMED Terminology Solutions

CAP Cancer Protocols: What Objects are We Modeling?

- Piece (s) of Tissue ?
- Surgical Specimen (s) ?
- Organ Specimen (s) ?
- Type of Neoplasm ?
- Diagnosis ?
- Blocks (s) ? Slides ? Aliquot (s) ?
- Special Tests ?
- Patient Findings ?
- Question/Answer Set / Pathology Synoptic Report
  - Leave the "object model" to the terminology

© 2007
College of American Pathologists
SNOMED Terminology Solutions

CAP Cancer Protocols: Where do we put essential metadata?

- Base objects (DEC, Value Domain, CDE, Value Items)
- Parent (Previous) objects (DEC, Value Domain, CDE, Value Items)
- Original objects (DEC, Value Domain, CDE, Value Items)
- Checklist Version
- Essential form presentation info
- Maps to other coding systems
- Essential algorithm support:
  - Q/A hierarchy and functionality
  - Skip areas (item dependencies): more complex than Form Builder approach
  - Template and sub-template injection
  - Calculations
  - Black box logic (stage calculations)
  - Validation
  - Repeating sections
  - Converging and diverging algorithmic pathways

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Need Better DEC Curation

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Selected Data Element

Public ID:	2431571
Version:	1.0
Long Name:	Cutaneous Melanoma TNM Finding Microscopic Positive Test Result Number Lymph Node Integer Count
Short Name:	2431365v1.0:2433291v1.0
Preferred Question Text:	Number containing lymph nodes involved microscopically
Definition:	definition pending_Too small to be seen except under a microscope.:A test result confirming the presence of a disease, condition, or microorganism. (NCI):A numeral or string of numerals expressing value, quantity, or identification.:Small, bean-shaped organs located along the channels of the lymphatic system. The lymph nodes store special cells that can trap bacteria or cancer cells traveling through the body in lymph. Clusters of lymph nodes are found in the underarms, groin, neck, chest, and abdomen. Also called lymph glands._A number with no fractional part._To determine the number or amount of something; the result of this activity.
Workflow Status:	RELEASED

Object Class

More Details

Public ID:	2431287
Version:	1.0
Long Name:	Cutaneous Melanoma TNM Finding
Short Name:	C48791
Context:	caBIG
Qualifier:	

Object Class Concepts

Concept Name	Concept Code	Public ID	Definition Source	EVS Source	Primary
Cutaneous Melanoma TNM Finding	C48791	2431171	NCI	NCI_CONCEPT_CODE	Yes

Property

Public ID:	2431234
Version:	1.0
Long Name:	Microscopic Positive Test Result Number Lymph Node
Short Name:	C25252:C35682:C25337:C12745
Context:	caBIG
Qualifier:	

Confusion between the roles of caDSR and terminologies like NCI Thesaurus and SNOMED CT

Property Concepts

Concept Name	Concept Code	Public ID	Definition Source	EVS Source	Primary
Microscopic	C25252	2204913	NCI	NCI_CONCEPT_CODE	No
Positive Test Result	C35682	2431156	NCI	NCI_CONCEPT_CODE	No
Number	C25337	2202701	Source => Name: NCI,	NCI_CONCEPT_CODE	No
Lymph Node	C12745	2202290	NCI-GLOSS	NCI_CONCEPT_CODE	Yes

What does it mean?
Where is the standardization?
How does this help?
Can this be used to compute concept subsumption or equivalence?

URU: Understandable, Reproducible, Useful: Fails on all three
There are many examples like this

© 2007
College of American Pathologists
SNOMED Terminology Solutions

An Approach to DEC/VD/CDE Creation

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Considerations for a SECCC Approach to the caDSR

- Do we need to go through UML, or can we just convert template→caDSR records?
- Protocol modeling must very generic, and must not attempt to model specific "neoplasm objects," or similar objects:
  - The generic modeled objects should be subtypes of "Protocol Question," not subtypes of "Neoplasm" or "Disease" or "Patient Finding," etc. We are modeling a synoptic report, NOT a tumor or a patient.
- Construct CDEs that will enable maximum query flexibility
  - Have sample queried in mind, and consider how the DSR will need to be searched
- Try to construct reusable, generic DEs and VDs whenever possible
- Avoid creating DEs specific to a single CAP Protocol – these are not reusable
- Create base classes (DE, VD, [CDE]) whenever possible, and derive Protocol-specific subclasses from them if absolutely required.
- Decide whether template metadata (e.g. control type, max value) belongs in the caDSR, or in an external template of some sort (e.g. XML, database)

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Non-Reusable CDE Styles

Template Version ID: Colon and Rectum Resection, Version 1.000.000
(Pointer to storage site of template metadata)

(Common) Data Element (CDE/DE)

Data Element Concept (DEC)

OBJECT: CAP Colon Carcinoma (Checklist)
Property: Histologic Type

DEC ID: 10001
Prev DEC ID: Null
Original DEC ID: 10001
Base DEC ID: 10001
Owner: CAP

SNOMED CID: 123456789
NCI_Meta_CUI:
LOINC:
Etc…

Specific and **not** reusable across Protocols

Value Domain (VD)

Rep Term: Enumeration
Colon Carcinoma Histologic Type Enum
ItemType: Single-Select
VD ID: 10001
Prev VD ID: Null
Original VD ID: 10001
Base VD ID: 10001
Owner: CAP

Requires specificity ("colon") in this case

Adenocarcinoma , other (specify type if known)
ItemType: AnswerChoice (Fill-in)

Value ID: 10001

? Prev Value ID: Null
? Original Value ID: 10001
? Base Value ID: 10001
Owner: CAP

SNOMED CID, NCI_Meta_CUI, LOINC, Numeric Value, Etc…

I don't favor this approach because the CAP Protocol Type (Colon) also must be specified by the selected Colon checklist template ID. Thus the word "Colon" is redundant in the Object Class, and the DEC is no longer reusable. It also causes unnecessary proliferation of DECs

Proposed CDE Style

Requires specificity ("colon") in this case

Data Element Concept (DEC)

OBJECT: CAP Cancer Protocol Question
Property: Histologic Type

DEC ID: 10001
Prev DEC ID: Null
Original DEC ID: 10001
Base DEC ID: 10001
Owner: CAP

SNOMED CID: 123456789
NCI_Meta_CUI:
LOINC:
Etc…

Generic and **reusable** in many Protocols

Value Domain (VD)

Rep Term: Enumeration
Colon Carcinoma Histologic Type Enum
ItemType: Single-Select
VD ID: 10001
Prev VD ID: Null
Original VD ID: 10001
Base VD ID: 10001
Owner: CAP

Adenocarcinoma , other (specify type if known)
ItemType: AnswerChoice (Fill-in)

Value ID: 10001

? Prev Value ID: Null
? Original Value ID: 10001
? Base Value ID: 10001
Owner: CAP

SNOMED CID, NCI_Meta_CUI, LOINC, Numeric Value, Etc…

Uses generic DECs and VDs whenever possible. These can be used as "Base DECs," "Base VDs" to derive generic (base) CDEs or more specific types when necessary. This will allow querying across multiple derived types, via the base or parent class.

© 2007
College of American Pathologists
SNOMED Terminology Solutions

To Clinical Annotation and BEYOND!!!

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Breakout Group 1: Vocab/Architecture Gurus

Modeling the Protocol Templates in UML/caDSR/Form Tool/EVS/SNOMED CT

- Can the template metadata (e.g. code mappings, inheritance fields) be represented in the current caDSR or other repositories? What modifications need to be made? Do we need to use a new repository for template metadata?
- How will the UML represent the hierarchical and inheritable/injectable template structure of the protocols?
  - Can the UML modeling step be bypassed, with direct creation of caDSR/EVS objects?
  - Can/Should the UML model be created via a direct XMI transform from the hierarchical template model?
  - Are there methods of programmatically creating DECs, Value Sets, CDEs from the template metadata? Is it necessary to got through UML for this first?
  - Are there available methods for directly importing the terminology (SNOMED CT), and creating or linking to EVS entries?
- What new standard terms do we need in the DSR to support the SECCC, e.g. Representation terms, categories, classifications, template versions, etc.

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Breakout Group 2: Clinical Use-Case Gurus

Developing a Framework for Clinical Datasets Beyond the Pathology Synoptic Reports (CAP Protocols)

- Many essential clinical oncology fields are not represented as minimal data sets or synoptic reports. We need specific and standardized synoptic questions with defined standard answer choices for each clinical area
  - Surgical oncology, Med/Peds/Gyn Oncology, Molecular diagnostics, Chemo protocols, related treatments, stem cell transplant (NMDP/ABMTR) forms, radiology, RadOnc, etc. Task: Add more groups and specific question sets as time permits. Prioritize the list.
  - Who will form the groups to produce these sets? Who will review them? Who will implement them in caBIG? Inter-Organizational structure?
  - Can these Q/A sets adopt the CAP Cancer Protocol template format, or is there a need for other formats? Are there other use cases that would require modifications to the SECCC template model?
  - What types of complex data are required for synoptic clinical annotation: do we really need x-rays, microarray data sets, path images, or is a summary adequate for clinical annotation (at this stage)? Is there a standard for radiology synoptic reporting? What standards are available for outcomes reporting and therapy selection in Oncology? Severity of Disease reporting? Adverse Reaction Scoring Systems? Cardiac Performance Status? Renal Status? Exercise Tolerance? Others?
  - Some other caBIG groups may be producing clinical annotations that could be "harvested" as synoptic Q/A pairs (e.g. microarray, radiology). Where do these exist, and how can these be converted to a standard Q/A structure appropriate for general clinical annotation and general synoptic reporting? Is there any caBIG organizational structure to bring this data into CAE etc? Can we implement a standard format for all forms of clinical annotation, similar to the SECCC?

© 2007
College of American Pathologists
SNOMED Terminology Solutions

Breakout Group 3: Question/Answer Sets (QAS) in Practice: Building the Tools

Critical need for user-friendly, fast tools: How can we organize to do this? CDC/PHIN Coordination? Priorities?

- **Distributed template editing**
  - Beyond the Forms tool – need to support more robust metadata-rich forms
  - Tools for creating, editing templates
  - Is there a pressing need to modify existing SECCC template structure at this point?
  - Must support hierarchical base templates, template inheritance, template injection, sub-template injection, algorithms, calculations
  - Dev language (.NET vs Java vs Access) to maximize contributors to project
  - How to connect .NET apps to existing Java APIs? caCore should support .NET or .NET must use a Java Bridge?
- **JIT Screen generator**
  - Based on above templates, need screen generator (web/thick forms/smart client) that annotates or plugs into other caBIG apps , as well as vendor and home-grown apps
  - Recommend an architecture and strategy based on the template model
    - Template source is web service, API, XML on public server??
    - App is web based, thick forms, smart client (click-once deployment) app??
    - Technology is roll-your-own (C#, Java)/AJAX, XAML, X-forms, PDF forms, InfoPath, ???
    - Data export formats, and direct connection to data repositories: formats?
- **Template repository**
  - How and Where will we store template versions: caDSR, Public XML, somewhere else??
- **Query generator tool**
  - Select template(s) and versions desired, and then query with generic or specific questions, using base-level CDEs when appropriate
  - Suggest architecture, SQL generation mechanisms, working with standard EAV-style data models through the Grid.

© 2007
College of American Pathologists
SNOMED Terminology Solutions

SECCC & caBIG
Divide and Conquer!

© 2007
College of American Pathologists
SNOMED Terminology Solutions